View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Proceedings Published by the LSA (Linguistic Society of America)

2017. *Proc Ling Soc Amer* 2, 41:1-15. https://doi.org/10.3765/plsa.v2i0.4097.

Who has more? The influence of linguistic form on quantity judgments

Gregory Scontras, Kathryn Davidson, Amy Rose Deal, & Sarah E. Murray*

Abstract. Quantity judgment tasks have been increasingly used within and across languages as a diagnostic for noun semantics. Overwhelmingly, results show that notionally atomic nouns (*Who has more **cats**?*) are counted, while notionally non-atomic nouns (*Who has more **milk**?*) are measured by volume. There are two primary outliers to the strict atomicity-tracking pattern. First, some nouns, like *furniture*, show primarily cardinality-based results in some studies, indicating atomicity, but nevertheless show systematic non-cardinality judgments in other studies, with comparison based instead on value/utility. Second, it has been reported that speakers of the Amazonian language Yudja favor cardinality-based quantity comparison for all nouns regardless of notional atomicity. In the current study, we show that both of these patterns arise in naïve English speakers in the absence of clear linguistic cues to atomicity, and suggest that the absence or mis-diagnosis of linguistic cues may be behind the reported outliers to atomicity-tracking.

Keywords. quantity judgments; atomicity; comparison strategies; informativity; mass-count distinction

1. Introduction. Quantity judgment tasks have been increasingly used within and across languages as a diagnostic for noun semantics, identifying the part-whole structure of the denotation of a nominal root. In classic work on these tasks, Barner & Snedeker (2005) demonstrate differential behavior in comparison judgments of the form *Who has more NOUN?*: where NOUN is an atomic noun (e.g., *cats*), experimental participants respond by counting, but where NOUN is non-atomic (e.g., *milk*), participants use a measurement strategy based on volume. Crucially, aggregate nouns like *furniture* pattern with other atomic nouns: participants perform quantity judgments by counting contextually-salient atoms. These patterns have lent crucial support to semantic theories arguing that nouns differ in the nature of their minimal parts—only some noun denotations have minimal parts (Bunt 1985), or alternatively, only some noun denotations have minimal parts that are stable (Chierchia 2010), not strongly connected (Grimm 2012), or not overlapping (Landman 2011). One consequence is that matters related to the mass-count distinction cannot be handled purely in terms of cumulativity, *pace* Chierchia (1998).

Recent research has uncovered two primary outliers to the strict atomicity-tracking pattern. A first challenge comes from English: certain aggregate nouns show cardinality-based judgments in some studies (Barner & Snedeker 2005), indicating atomicity, but systematic non-cardinality judgments in other studies, with comparison based instead on value/utility (Grimm & Levin 2012). A second challenge comes from the growing crosslinguistic quantity judgment literature: Lima (2014) reports that speakers of Yudja (Tupí; Brazil) favor cardinality-based quantity comparison

*We would like to thank Angelika Kratzer, Manfred Krifka, Jesse Snedeker, Beth Levin, members of the 2015-2016 SIAS summer institute on The Investigation of Linguistic Meaning, audience members at the 2017 LSA Annual Meeting in Austin, the Wissenschaftskolleg zu Berlin, and the National Humanities Center. Authors: Gregory Scontras, University of California, Irvine (g.scontras@uci.edu), Kathryn Davidson, Harvard University (kathryndavidson@fas.harvard.edu), Amy Rose Deal, University of California, Berkeley (ardeal@berkeley.edu), & Sarah E. Murray, Cornell University (sarah.murray@cornell.edu).

for all nouns, including those that name non-atomic substances. In this paper, we propose that these behaviors correlate with the presence or absence of clear linguistic cues to atomicity in the relevant environments, and we demonstrate that similar patterns can be produced experimentally precisely by taking linguistic cues away. In performing a quantity judgment, the linguistic form (i.e., cues to atomicity) is the primary driver of behavior (i.e., counting vs. measuring). In the absence of linguistic cues, behavior is influenced by the salience of 1) concrete portions and 2) alternative dimensions of measurement, all in an attempt on the part of participants to provide informative answers to the quantity judgment prompt. This suggests that the outlier findings need not perturb the general argument that quantity judgments track the atomicity of noun denotations.

2. Background: Exceptions to the atomicity-tracking pattern. Here we review two recently-uncovered exceptions to the strict atomicity-tracking pattern of quantity judgments.

2.1 GRIMM & LEVIN (2012). Barner & Snedeker (2005) found striking evidence for the countability of aggregate nouns like *silverware* (their "object-mass" nouns). However, Grimm & Levin (2012) raise serious concerns about the stimuli used to elicit counting behavior for these nouns. Specifically, Grimm & Levin argue that the depicted scenes ignored heterogeneity, a central property of aggregate nouns. Unlike mass nouns, which denote homogeneous substances (e.g., toothpaste, water, milk, etc.), Grimm & Levin argue that aggregate nouns like *furniture* denote heterogeneous collections of entities (e.g., chairs, tables, desks, etc.). We see an example stimulus from Barner & Snedeker (2005) for the aggregate noun *silverware* in the left panel of Fig. 1: all of the pictured objects are of the same type (i.e., forks), differing only in size.

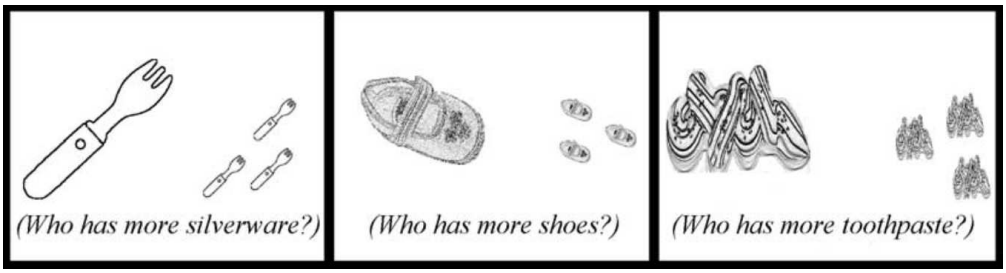

Figure 1: Example stimuli from Barner & Snedeker (2005).

According to Grimm & Levin, heterogeneity emerges through the canonical events associated with aggregate nouns (e.g., furnishing a space in the case of *furniture*). Because these nouns are necessarily associated with events, Grimm & Levin argue that comparison may target the fulfillment of the function of these events (e.g., how well a given collection of furniture would furnish a space). Indeed, they find evidence of this claim in a modified version of the quantity judgment task featuring heterogeneous stimuli for aggregate nouns. In this modified task, participants demonstrated a clear preference for function (e.g., a sofa, an easy chair, a coffee table, and a small bookcase) over cardinality (e.g., one table and four chairs) in their comparison judgments.

Grimm & Levin (2016) build on these results, proposing that *furniture*-type aggregate nouns are in fact non-countable because their associated events do not restrict the counting domain to single entities. Thus, the authors use the seemingly non-atomic behavior in the quantity judgment task to motivate a new semantics for aggregate nouns.

2.2 LIMA (2014). The other notable counterexample to the strict atomicity-tracking pattern in quantity judgments comes from Lima (2014), who presents data from the language Yudja. Yudja is an Amazonian language of the Tupí family spoken by approximately 300 people in the Xingú Indigenous Park, Matto Grosso, Brazil. Lima tested 18 native speakers of Yudja in the quantity judgment task, looking at notional count (e.g., 'bowl,' 'spoon'), aggregate (e.g., 'clothes,' 'ceramic'), and notional mass (e.g., 'flour,' 'water') nouns. As expected, participants reliably counted when performing quantity judgments for notional count and aggregate nouns. However, participants counted to the *same* extent for notional mass nouns.

If counting behavior in the quantity judgment task signals an atomic nominal denotation, then *y'a* 'water' in Yudja is just as atomic as *karaxu* 'spoon.' Indeed, this is precisely the proposal that Lima advances. She proposes that all nouns in Yudja refer at the kind level, but instantiate as maximally self-connected instances of the kind in context (cf. the framework of mereotopology from Grimm 2012). Thus, like Grimm & Levin (2012, 2016), Lima uses nonstandard behavior in the quantity judgment task to motivate a nonstandard analysis of nominal semantics.

3. Cross-linguistic considerations. We have so far considered two counterexamples to the atomicity-tracking pattern of behavior observed by Barner & Snedeker (2005) for quantity judgment tasks: 1) ostensibly atomic aggregate nouns like *furniture* getting measured instead of counted (Grimm & Levin 2012), and 2) notional substance nouns like *y'a* 'water' in Yudja getting counted instead of measured (Lima 2014). Before pursuing our follow-up investigation of these apparent exceptions to the atomicity-tracking behavior, we believe it is worthwhile to pause and note that languages differ with respect to how they express morphosyntactic cues to atomicity. Given the central role that these cues play in the quantity judgment task, it is imperative to understand precisely which cues are present in the task prompt.

In English, noun words come fully specified for the atomicity of their denotation.[1] Barner & Snedeker (2005) demonstrate this fact using nouns like *rope* that, depending on their morphosyntax, optionally refer to atomic individuals or non-atomic substances. In (1-a), *ropes* appears with atomic morphosyntax, and participants use this cue to guide them toward counting in their responses to the quantity judgment prompt. In (1-b), *rope* appears with non-atomic morphosyntax, and participants use this cue to guide them toward measuring volume. Given that nouns in English come specified for atomicity, whenever a noun appears in the quantity judgment prompt, so too do morphosyntactic cues to atomicity. The only way to remove these cues (in English) is to remove the nouns from the prompt, as in (1-c). We follow up on this idea in the experiments presented below.

(1) a. Who has more ropes? [*atomic*]
b. Who has more rope? [*non-atomic*]
c. Who has more?

Crucially, not every language is like English. That is, not every language provides its morphosyntactic cues to atomicity on the noun in the quantity judgment prompt. In Cheyenne, an Algonquian language spoken in Montana and Oklahoma, those cues appear on the verb, (2): comparison is expressed with the verbal prefix *hehpe-* and the verb has to specify either quantity or size.

[1]The contrast here is with noun *roots*, where the issue is subject to more debate.

In Swedish, morphosyntactic cues to atomicity appear on the quantifier, as in (3).[2] In either language, omitting the noun from the quantity judgment prompt does not eliminate the morphosyntactic cues to atomicity. Indeed, any version of the quantity judgment prompt will necessarily include cues to atomicity, and therefore clearly signal a comparison strategy to experimental participants.

(2) *Cheyenne*[3]

a. Taaso tse-hehpėstoha?
which IND-be.more.in.quantity
'Which is more (in number)?' [*atomic*]

b. Nevaaso tse-ho'tsė-stse tse-hehpėstȯha-tse?
who IND-have.s.t.-CNJ.3SG IND-be.more.in.quantity-CNJ.REL
'Who has more (in number)?' [*atomic*]

c. Taaso tse-hehpao'o?
which IND-be.more.in.size
'Which is more (in size)?' [*non-atomic*]

(3) *Swedish*[4]

a. Vem har flest katter?
who has most cats
'Who has more cats?' [*atomic*]

b. Vem har mest vatten?
who has most water
'Who has more water?' [*non-atomic*]

In Nez Perce, the trend goes in the opposite direction: even nouns typically lack clear morphosyntactic cues to atomicity (Deal 2016); only adjectives reliably host this morphology. Thus, in Nez Perce, the quantity judgment prompt generally contains atomicity cues only when the noun appears modified by an adjective (Deal 2017).

[2]This of course is at least theoretically the case in English, too, where *fewer* supposedly contrasts with *less* in precisely a parallel way. We find that many English speakers, ourselves included, do not reliably distinguish *fewer* from *less*. In a pilot study, nine English-speaking participants on Amazon.com's Mechanical Turk did not show any significant differences in counting/measuring responses to *Who has fewer?* versus *Who has less?*.

[3]Cheyenne examples are from Sarah Murray's fieldwork. Orthography: dots over vowels indicate voicelessness; ' is a glottal stop. Glosses: CNJ dependent clause (conjunct) suffix, IND indicative dependent clause (conjunct) prefix, REL relational, SG singular. See also Snider & Murray (to appear) on comparison in Cheyenne.

[4]Swedish examples are from Filippa Lindahl, personal communication. Swedish distinguishes mass and count quantifiers in positive, comparative, and superlative grades (*mycket/mer/mest* 'much/more$_{mass}$/most$_{mass}$' vs. *många/flera/flest* 'many/more$_{count}$/most$_{count}$'); superlative forms are judged most appropriate for the quantity judgment question.

(4) *Nez Perce*[5]

a. 'Isii-nm 'uus qetu 'ilexni?
who-GEN has COMP A.LOT
'Who has more?'

b. 'Isii-nm 'uus qetu 'ilexni teewtes?
who-GEN has COMP A.LOT rope
'Who has more rope(s)?'

c. 'Isii-nm 'uus qetu 'ilexni ti-ta'c teewtes?
who-GEN has COMP A.LOT PL-good rope
'Who has more good ropes?' [*atomic*]

d. 'Isii-nm 'uus qetu 'ilexni ta'c teewtes?
who-GEN has COMP A.LOT good rope
'Who has more good rope?' [*non-atomic*]

It bears noting that the two exceptions that have been observed to atomicity tracking in the quantity judgment task involve cases without clear morphosyntactic cues to atomicity. In Yudja, all nouns lack atomicity cues. In English, one could make the case that aggregate nouns lack clear atomicity cues. In both cases, experimenters have observed seemingly atypical quantity judgment behavior. It seems plausible that this behavior might arise when clear cues to atomicity are absent from the quantity judgment prompt. To investigate this hypothesis, one must establish a baseline of behavior: what happens in the absence of atomicity cues? To that end, we present two experiments testing the quantity judgment behavior of naïve speakers of English when these cues are absent.

4. Experiment 1: Removing cues. Our first experiment investigates how speakers perform quantity judgment tasks in the absence of clear linguistic cues to atomicity. Taking advantage of properties of English, we remove those cues by removing altogether the noun from the quantity judgment prompt. To establish a baseline against which to compare these results, we contrast responses to the noun-less prompt with responses to a prompt that contains the relevant nouns.

4.1 PARTICIPANTS. We recruited 45 participants via Amazon.com's Mechanical Turk crowdsourcing service. Participants were compensated for their participation. Based on self-reports, 43 participants were identified as native speakers of English; their data were included in the analyses reported below.

4.2 DESIGN AND METHODS. Participants were tasked with performing quantity judgments for a variety of scenes. Each scene featured an image with the relevant material appearing on both the right and left sides of the image. Participants were told that "whatever is on the left of the scene belongs to one person, and whatever is on the right belongs to a different person." Their job was to make judgments about the scenes and to write a few words about why they made the judgments they did (Fig. 2). We included these justifications so that we could reconstruct the reasoning behind the judgments performed in the absence of linguistic cues to atomicity.

Scenes were constructed so that the material on one side was greater by cardinality, while the other side was greater by volume (e.g., three small puddles of milk vs. one large puddle in Fig. 2). Scenes featured referents from one of two ontological categories: either individuals (denoted by

[5]Nez Perce examples are from Amy Rose Deal's fieldwork. See Deal (2017) for discussion.

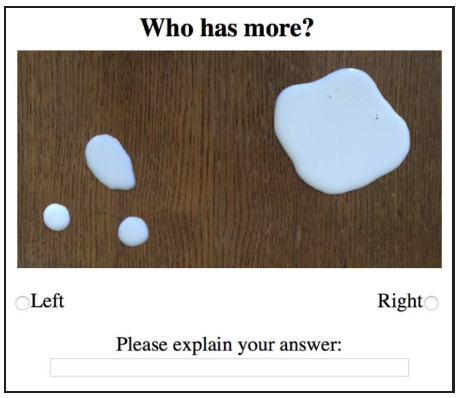

Figure 2: Example `no-noun` trial from Expt. 1; the left side has more milk by cardinality (counting puddles) while the right side has more milk by volume.

count nouns) or substances (denoted by mass nouns). The full list of test items appears in Table 1. We included a single catch trial, *eggs*, in which cardinality and volume favored the same side (i.e., one egg vs. three eggs). Responses were coded according to whether participants chose the side with greater cardinality, indicating that they had performed their quantity judgment via counting.

individual	**substance**
cups	dirt
flowers	fabric
jars	flour
knives	milk
leaves	paper
mugs	sugar
rocks	water
socks	
spoons	

Table 1: Test items from Expt. 1.

Between subjects, we manipulated the quantity judgment prompt. Participants were randomly assigned to either the `no-noun` (n=17) or the `two-noun` (n=26) condition. In the `no-noun` condition, participants were asked "Who has more?" Thus, in the `no-noun` condition, participants received no linguistic cues to atomicity in the quantity judgment prompt. In the `two-noun` condition, participants were asked "Who has more NOUN?" with the relevant noun, either bare mass or plural count, appearing in the prompt.

Participants completed a total of 17 trials (16 test items and the single *eggs* catch trial). Items were presented in a random order for each participant. After testing, participants completed a short demographics questionnaire in which they indicated their native language; we analyzed the data from those participants who indicated that their native language was English.

4.3 RESULTS. For our *eggs* catch trial, 100% of participants chose the side with greater cardinality (and greater volume), indicating that they were attending to the experiment.

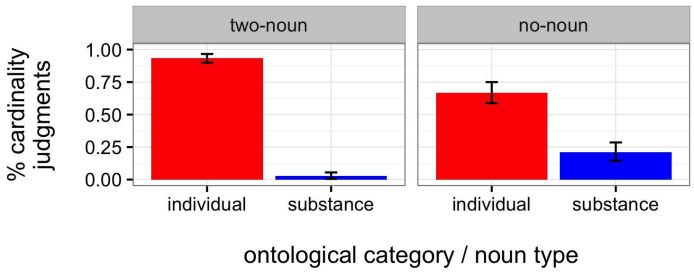

Figure 3: Results from Expt. 1.

Results from the 16 test items appear in Fig. 3, which plots rates of cardinality choices by ontological category and experimental condition. Visual inspection of the plot reveals categorical behavior for the `two-noun` condition: if participants saw an individual-denoting count noun, they performed a quantity judgment based on cardinality; for substance-denoting mass nouns, participants based their judgments on volume. In the `no-noun` condition, we observe the same general tendency, but with seemingly less categorical behavior.

To confirm these visual trends in the data, we analyzed participants' responses using a mixed-effects logistic regression model predicting cardinality choices with fixed effects of ontological category (`individual` vs. `substance`), experiment condition (`two-noun` vs. `no-noun`), together with their interaction. The model included random intercepts for participants and items (i.e., the maximal random effects structure justified by the data). The model revealed a main effect of ontological category (β = -9.03, z = -7.86, $p < 0.01$): scenes depicting non-atomic substances received fewer quantity judgments based on cardinality than scenes depicting atomic individuals. The model also revealed a significant interaction between ontological category and experiment condition (β = 5.63, z = 4.43, $p < 0.01$): in the `no-noun` condition, results were indeed less categorical, deviating from the canonical atomicity-tracking baseline.

Having confirmed the initial hypothesis that behavior in the quantity judgment task becomes less categorical in the absence of clear linguistic cues to atomicity, we next set out to understand the factors that characterize the deviant behavior. To that end, we looked at participants' justifications for trials in which they failed to provide the canonical judgment.

Given that behavior was largely categorical in the `two-noun` condition, there are very few cases of non-canonical quantity judgments. Of the 260 responses to individual-denoting nouns in the `two-noun` condition, 20 (8%) were performed on the basis of volume, rather than cardinality. Eight of these volume-based responses came from a single participant who justified every judgment using "surface area" (e.g., "the right spoon has more surface area"). Five other seemingly-volume-based judgments were provided in response to the scene with flowers, which some participants interpreted as depicting multiple flowers on both sides (Fig. 4); given the ambiguity in the image, we hesitate to over-interpret these responses. The remaining 7 volume-based judgments came from 5 separate participants and were given in response to "cups," "mugs," "rocks," and "socks."

Turning to substance-denoting nouns in the `two-noun` condition, just 5 of the 182 responses (3%) were judgments based on cardinality, rather than volume. These responses came

Figure 4: The "flowers" scene from Expt. 1; some participants viewed the scene as depicting multiple flowers on both sides.

from two participants who used quantizing nouns to partition the measured substance (e.g., *piles*, *drops*, *pieces*, or *lines*), then counted the partitions in performing the comparison.

Next, we examined justifications to seemingly non-canonical behavior in the `no-noun` condition: volume-based judgments for scenes depicting atomic individuals, and cardinality-based judgments for scenes depicting non-atomic substances. Of the 170 responses to individual-depicting scenes in the `no-noun` condition, 56 (33%) were judgments performed on the basis of volume, rather than cardinality. There were 11 responses to the flowers scene, which we ignore given the concerns about the visual display that were described above. We also hesitate to over-interpret the 10 volume-based responses to the rocks scene. The noun *rock* switches between count and mass usage (Barner & Snedeker 2005): depending on the morphosyntactic cues to atomicity expressed, the noun can either reference atomic individuals (e.g., *many rocks*) or non-atomic substances (e.g., *much rock*). Without seeing the intended form of the noun in the `no-noun` quantity judgment prompt, participants were free to imagine whichever form they preferred and perform the judgment on the basis of that form.

That leaves 35 judgments for atomic individuals performed on the basis of volume. The majority of these judgments, 21, were justified in terms of size (e.g., "bigger," "longer," etc.). Interestingly, these 21 judgments came from just four participants; these participants consistently demonstrated measuring (as opposed to counting) behavior in the task. There were 8 judgments justified on the basis of capacity (e.g., "the single jar can hold more," "fit more in," etc.). The remaining 6 volume-based judgments were evenly split between value justifications (e.g., "big knives are more expensive," "larger knife is better") and utility justifications (e.g., "bigger jars are more useful," "this spoon is more useful").

Turning to substance-depicting scenes in the `no-noun` condition, of the 119 responses, 25 (21%) were quantity judgments performed on the basis of cardinality. These judgments all came from just four participants who consistently counted as they performed their quantity comparisons. Looking at the justifications, 22 of the cardinality-based judgments were justified in terms of counting a partitioned substance using a quantizing noun (e.g., "there are more pieces," "right side has 3 piles versus one," "three droplets on the right and only one on the left," etc.). The remaining 3 cardinality-based judgments were justified by counting alone (e.g., "right has 2, left has 1," "left

has 4, right has 1," etc.).

4.4 DISCUSSION. Our results replicate the finding from the previous literature that quantity judgments in the presence of clear morphosyntactic cues to atomicity are near-categorical, with atomic individuals judged on the basis of cardinality (92% of the time) and non-atomic substances judged on the basis of volume (97% of the time). In the absence of atomicity cues, as in our `no-noun` condition, judgments still follow the general atomicity-tracking trend, but deviate significantly from the categorical baseline. Based on the justifications participants provided, clear strategies emerge. When it comes to counting non-atomic substances, participants overwhelmingly partitioned the substance using quantizing nouns (i.e., "atomizers"; cf. Scontras 2014) and then counted members of the partitioned denotation. To compare the volume of atomic individuals, participants focused on alternative dimensions of measurement (e.g., value, utility, size) and performed their comparison along those dimensions.

Both of these strategies ring familiar from the experimental literature, especially the notable exceptions mentioned in Section 2. Recall that Grimm & Levin (2012) found measuring behavior for atomic aggregate nouns like *furniture* or *jewelry*, with participants justifying their responses in terms of relative utility. We found the same behavior with run-of-the-mill (and homogeneous) sets of objects like knives, suggesting Grimm & Levin's explanation based on the elaborate denotations of aggregate nouns might not be on the right track. We offer an alternative interpretation of the behavior Grimm & Levin observed: without clear morphosyntactic cues to atomicity, participants are free to perform their quantity judgments on the basis of whichever dimension they find most salient, or relevant to the task at hand. Participants recognize that the relative cardinalities of the competing sets are clear in context; why would a speaker ask after something so obvious? In an attempt to charitably interpret an otherwise trivial question and thereby provide an informative answer, some participants consider dimensions other than cardinality as they perform their quantity judgments. We return to this point in Section 6 below.

We turn next to the observed substance-counting behavior, whereby participants partitioned an otherwise non-atomic substance to deliver appropriate atoms for counting. This strategy appears reminiscent of the Yudja results from Lima (2014), who found near-total counting behavior regardless of the noun. In principle, one might have interpreted Lima's results as evidence for a default counting strategy in the absence of clear linguistic cues to atomicity—perhaps the lack of atomicity cues in Yudja led to the default counting strategy. However, our results argue against this interpretation. We found that counting partitioned substances is by no means the default strategy in the absence of atomicity cues: just 21% of the responses—and only four participants—employed this partition-counting strategy. It would seem, then, that partition-counting is a viable but by no means obligatory strategy when performing quantity judgments in the absence of linguistic cues to atomicity.

Having found that judgments in the quantity judgment task deviate from the canonical atomicity-tracking pattern in the absence of clear morphosyntactic cues, the question then shifts to how participants settle on the comparison strategy they use without the cues present. A very plausible hypothesis holds that participants fill in these cues mentally, imagining an inflected noun as they perform the quantity judgment task. This imagined noun would then serve the same role it does when it appears overtly in the quantity judgment prompt, delivering canonical behavior. We test

this hypothesis in the following experiment.

5. Experiment 2: Filling in cues. Our second experiment follows up on the results of Expt. 1 by testing the hypothesis that speakers are mentally filling in linguistic cues to atomicity in the quantity judgment task when those cues are absent from the task prompt. We repeat the `no-noun` and `two-noun` conditions from Expt. 1, and to them add two new conditions which 1) explicitly invite participants to fill in a noun in the quantity judgment prompt before performing the comparisong, and 2) elicit responses to a broader set of nouns in the judgment prompt (e.g., partitioning and singular count nouns).

5.1 PARTICIPANTS. We recruited 254 participants via Amazon.com's Mechanical Turk crowdsourcing service. Participants were compensated for their participation. Based on self-reports, 253 participants were identified as native speakers of English; their data were included in the analyses reported below.

5.2 DESIGN AND METHODS. The design was similar to that of Expt. 1: participants saw the same context story, viewed images, performed quantity judgments, and provided short justifications for their judgments. We repeated the `no-noun` and `two-noun` conditions from Expt. 1. To these we added a new `fill-in` condition in which participants filled in missing material in the quantity judgment question prompt before performing the judgment and providing a justification; an example `fill-in` trial appears in Fig. 5. Filled-in material was coded according to whether it was bare mass (e.g., *milk*), partitioned mass (e.g., *puddles of milk*), plural count (e.g., *knives*), or singular count (e.g., *knife*). In order to compare responses to the filled-in material against a baseline with responses that explicitly provided that material, we also added a new `four-noun` condition in which participants encountered the "Who has more NOUN?" prompt with a broader set of nouns: either bare mass, partitioned mass, plural count, or singular count. We drew the partitioning quantizing nouns (e.g., *puddles*, *pieces*, etc.) from the most frequent partitioning justifications that participants provided for the relevant items in Expt. 1.

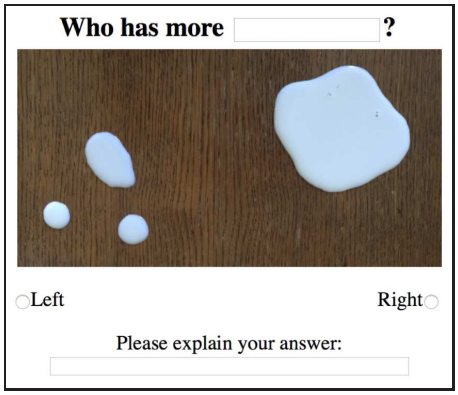

Figure 5: Example `fill-in` trial from Expt. 2.

Participants were randomly assigned to one of the four conditions: `no-noun` (n=63), `two-noun` (n=68), `four-noun` (n=63), and `fill-in` (n=59). We included a subset of the images used in Expt. 1: we removed the *flowers* trial, given the concerns described in the discussion above. Participants completed a total of 16 trials (15 test items and the single *eggs* catch trial). Items were

presented in a random order for each participant. Judgment responses were coded according to whether they favored the side with greater cardinality. After testing, participants completed a short demographics questionnaire in which they indicated their native language.

5.3 RESULTS. For the `fill-in` condition, some participants failed to follow instructions and instead of filling in a noun to complete the quantity judgment prompt, they provided their answer to the noun-less prompt (i.e., "right" or "left") in the relevant text field. We removed 435 of these incorrect responses, 11% of the total 4048 responses. Looking at the remaining responses, 99% of participants correctly answered the *eggs* catch trial, signaling that participants were indeed paying attention to the task.

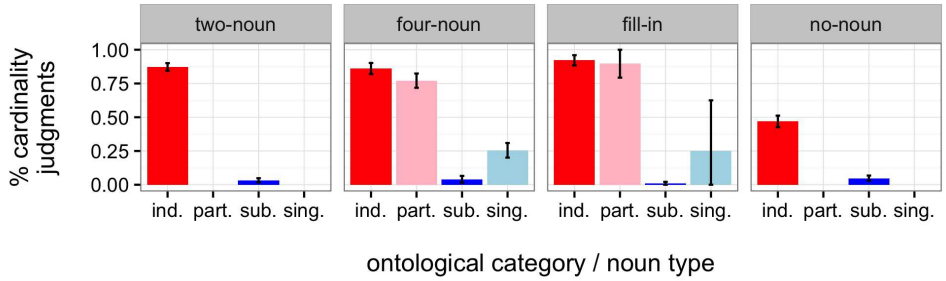

Figure 6: Results from Expt. 2. Error bars represent bootstrapped 95% confidence intervals.

Results from the 15 test items appear in Fig. 6, which plots rates of cardinality choices by experimental condition and ontological category/noun type. Visual inspection of the plot reveals a replication of the categorical behavior in the `two-noun` condition from Expt. 1, as well as deviation from the categorical baseline in the `no-noun` condition. In the current experiment, it also appears to be the case that the `no-noun` condition received fewer cardinality ratings overall. Turning to the `four-noun` and `fill-in` conditions, individual-denoting plural count or partitioned mass nouns deliver counting behavior, while substance-denoting bare mass or singular count nouns deliver volume-based measuring behavior.[6] Crucially, it appears that participants are providing similar quantity judgments regardless of whether they received the noun in the prompt (`two-noun`, `four-noun`) or they provided the noun themselves (`fill-in`). However, the `no-noun` condition seems to stand apart from the other three in its pattern of responses.

To confirm these visual trends, we began by analyzing the replication of Expt. 1's conditions, comparing cardinality choice rates in the `two-noun` and `no-noun` conditions. We fit a mixed effects logistic regression model predicting cardinality choices with fixed effects of ontological category (`individual` vs. `substance`) and condition (`two-noun` vs. `no-noun`), together with their interaction. The model included random intercepts for participants and items (i.e., the maximal random effects structure justified by the data). The model revealed a main effect of ontological category (β = -6.69, z = -11.24, $p < 0.01$), as well as a significant interaction between ontological category and condition (β = 4.66, z = 7.28, $p < 0.01$). These results replicate the findings of Expt. 1, where responses are much more categorical in the `two-noun` vs. the `no-noun` condition, with responses to both conditions driven largely by ontological category. However, un-

[6]We received only 8 singular count noun responses in the `fill-in` condition, which is why the confidence interval appears so large in Fig. 6.

like Expt. 1, the model also finds a main effect of condition (β = -1.54, z = -3.45, $p < 0.01$): there were fewer cardinality choices overall in the `no-noun` condition.[7]

Next, we compared the `fill-in` condition with the other three conditions. We fit a mixed effects logistic regression model predicting cardinality choices with the fixed effect of condition (`two-noun`, `four-noun`, `fill-in`, `no-noun`); we dummy-coded the condition predictor, with `fill-in` as the reference level. The model included random intercepts for participants and items (i.e., the maximal random effects structure justified by the data). The model revealed a main effect of condition for the `no-noun` contrast (β = -1.33, z = -4.44, $p < 0.01$); responses to the `no-noun` condition differed significantly from responses to the `fill-in` condition. The other two contrasts were not significant (`two-noun`: β = 0.19, z = 0.66, $p > 0.50$; `four-noun`: β = 0.28, z = 0.96, $p > 0.33$), signaling that responses to the `fill-in` condition did not differ significantly from the `two-noun` or `four-noun` conditions.

5.4 DISCUSSION. We continue to find that the main driver of comparison behavior in the quantity-judgment task is the linguistic form of the noun used in the task prompt. We see near-categorical behavior in the `two-noun` and `four-noun` conditions, where participants encountered a fully-inflected noun. Interestingly, in the `four-noun` condition, the noun given overrode whatever preference might have been introduced by the ontological category of the materials pictured: non-atomic substances were counted if they were named with a partitioning noun (e.g., *pieces*, *drops*) while atomic individuals were measured by volume if they were named with a "massified" singular count noun (e.g., *knife*, *leaf*). Crucially, we found no measurable difference between the conditions where participants were given a noun explicitly in the judgment prompt (i.e., `two-noun` and `four-noun`) and the `fill-in` condition where participants provided that noun themselves. In all cases, task behavior was determined by the noun.

However, behavior in the `no-noun` condition clearly stands apart, leaving open the question of precisely how speakers perform the task in this condition. Given the marked difference between the `no-noun` and `fill-in` conditions, we lack support for our hypothesis that participants are simply (mentally) filling in a noun as they perform quantity judgments in the `no-noun` condition. Whatever participants *are* doing, it does not align with the strategy from the `fill-in` condition. Of course, there remains the possibility that participants are still filling in material in the `no-noun` condition prompt, just not the same material they provide in the `fill-in` condition. Unfortunately, this revised hypothesis runs the risk of being unprovable: what sort of silent material gets filled in, and how could we possibly hope to know?

We are also left wondering why participants in this iteration of the `no-noun` condition provided fewer cardinality judgments than in Expt. 1's `no-noun` condition. For now we can only

[7] We confirmed this difference between Expt. 1 and Expt. 2 in an explicit comparison between the two experiments. We fit a mixed-effects logistic regression model predicting cardinality choices with fixed effects of ontological category (`individual` vs. `substance`), condition (`two-noun` vs. `no-noun`), and experiment (`expt1` vs. `expt2`). The model also included all possible two-way interactions between the fixed-effect predictors, as well as random intercepts for participants and items. The model revealed main effects of ontological category (β = -7.16, z = -10.70, $p < 0.01$) and condition (β = -1.29, z = -3.15, $p < 0.01$), as well as a significant interaction between ontological category and condition (β = 4.99, z = 8.88, $p < 0.01$). The model also revealed a main effect of experiment (β = -1.24, z = -2.63, $p < 0.01$), signaling that Expt. 2 received fewer cardinality-based choices overall, driven largely by the lower rates of cardinality choice in Expt. 2's `no-noun` condition.

speculate that the large difference in sample sizes between the two experiments (17 participants in Expt. 1 vs. 63 in Expt. 2) was the primary driver of the differences, and that the condition averages from Expt. 2 are more representative of the population behavior. A look at the justifications from the two sets of responses reveals similar strategies in both.

6. General discussion. Taken together, our results demonstrate the central role of morphosyntactic cues to atomicity in quantity judgment tasks (cf. Barner & Snedeker 2005). Confronted with linguistically-cued atomic nominals (i.e., plural count or partitioned mass nouns), participants perform quantity judgments by counting the named atoms; for linguistically-cued non-atomic nominals (i.e., bare mass and singular count), participants perform quantity judgments by measuring the volume of the relevant substances. Thus, the primary driver of behavior in the quantity judgment task is the quantity judgment prompt itself. When the prompt fails to provide the information necessary to determine atomicity, as in our `no-noun` conditions, participants default to the ontological category of the visualized material (i.e., atomic individuals vs. non-atomic substance). However, a substantial number of responses avoided this default atomicity-tracking behavior in the `no-noun` condition, instead counting contextually-salient atoms for the non-atomic substances and, more often, measuring atomic individuals along some alternative dimension (e.g., value or utility).

We attempted to gain some clarity on precisely how participants arrived at the strategies they used in the `no-noun` condition by directly comparing that condition with a condition in which participants explicitly filled a noun in before performing comparisons. The thought was that in the `no-noun` condition, participants were simply imagining a noun and performing their comparisons on the basis of the atomicity of the imagined noun. Thus, the `no-noun` prompt was viewed as an ellipsis structure where the elided phrase determined the compared material. However, the striking differences between this `fill-in` condition and the `no-noun` condition suggest that our hypothesis was misguided: whatever participants are doing in the `no-noun` condition, it does not appear to be the same as what they are doing when they explicitly fill in a noun in the quantity judgment prompt.

It would seem that the `no-noun` quantity judgment prompt is truly underspecified with regard to the atomicity of the compared material. With an underspecified quantity judgment prompt, participants must rely on contextual cues to perform the task. In the case of our experiment, context privileged various pieces of information. In scenes depicting atomic individuals, particularly salient was the differential in cardinality between the two sides of the display: three is clearly greater than one. If participants recognized the salience of this fact, then they could have reasoned backward from the perspective of the speaker asking the *who has more?* question: clearly the speaker can identify the cardinality differential, so pointing this out to him/her would not be very informative, useful, or cooperative. Thus, in an attempt to interpret the underspecified question as truly information-seeking, participants sometimes provided a response that itself could be informative (i.e., not immediately apparent from the scene). For atomic individuals, these purposefully-informative answers draw on alternative dimensions of measurement (i.e., size, value, or utility; cf. Grimm & Levin 2012).

In scenes depicting non-atomic substances, presumably the same pressures toward informativity apply. However, our results suggest that non-canonical responses (i.e., responses that do

not track atomicity) are much less common for substances, which might therefore suggest that the volume differential was less obvious (i.e., more informative) in our visual displays. Put differently, the difference in volume apparent in our visual stimuli might have been less salient than the difference in cardinality, which is why participants commented on the volume differential *more* in their responses to the underspecified prompt.

However, the categorical response patterns observed for the `two-noun` and `four-noun` conditions suggest that all of these considerations can reasonably be ignored once the judgment prompt itself establishes the atomicity of the compared material with morphosyntactic cues. Whatever the source of the non-canonical patterns, it bears noting that unlike the "count-everything" strategy observed by Lima (2014) for Yudja, if anything it would appear that the tendency in English goes in the opposite direction: measure everything. We have no evidence that counting salient atoms is the default strategy, which brings back into the focus the novelty of the Yudja results, and calls into question whether the Yudja prompt might have mandated counting behavior. According to Lima, the semantics of Yudja nouns are atomic; so, she claims, it is the always-atomic nominal in the quantity judgment prompt that delivers counting behavior. But what if other aspects of the linguistic prompt were to blame?

Indeed, we saw that languages differ with respect to their morphosyntactic expression of atomicity in the quantity judgment prompt. Some languages express these cues on the noun (e.g., English); others express them on the quantifier (e.g., Swedish) or even the verb (e.g., Cheyenne); yet others express these cues only on attributive adjectives (e.g., Nez Perce). Before one can use the quantity judgment task as a diagnostic for fine-grained details of nominal semantics in a language, it is imperative to clarify exactly how the morphosyntax of that language expresses cues to atomicity in the prompt itself. The current work serves to demonstrate the various strategies that arise when these cues are not fully specified.

7. Conclusion. In the absence of clear linguistic cues to atomicity, participants are not simply filling these cues in (cf. `no-noun` vs. `fill-in` in Expt. 2); rather, they draw on properties of the task context in an attempt to provide informative answers to the quantity judgment prompt. In cases where the relative cardinality obviously differed (i.e., 3 is obviously greater than 1), participants interpreted the question broadly. Our results indicate that such quantity judgments are subject to a variety of factors, including the contextual availability of salient portions and alternative dimensions of measurement. However, the primacy of linguistic form in determining behavior in this task demonstrates the importance of understanding the cues to atomicity provided by the quantity judgment prompt in a given language (e.g., Yudja). Cross-linguistic work using quantity judgments to diagnose nominal semantics should consider these potentially-confounding factors in the interpretation of comparison behavior, especially given that languages differ with respect to their expression of morphosyntactic cues to atomicity (e.g., plural marking, quantifier choice).

References

Barner, David & Jesse Snedeker. 2005. Quantity judgments and individuation: Evidence that mass nouns count. *Cognition* 97. 41–66.

Bunt, Harry C. 1985. *Mass terms and model-theoretic semantics*. Cambridge: Cambridge University Press.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6. 339–405.

Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174. 99–149.

Deal, Amy Rose. 2016. Plural exponence in the Nez Perce DP: A DM analysis. *Morphology* 26. 313–339.

Deal, Amy Rose. 2017. Countability distinctions and semantic variation. *Natural Language Semantics* 25. 125–171.

Grimm, Scott. 2012. *Number and individuation*. Stanford, CA: Stanford University dissertation.

Grimm, Scott & Beth Levin. 2012. *Who has more furniture?* An exploration of the bases for comparison. Mass/Count in Linguistics, Philosophy and Cognitive Science Conference, École Normale Supérieure, Paris, France, December 20–21, 2012.

Grimm, Scott & Beth Levin. 2016. Artifact nouns: Reference and countability. 42nd Annual Meeting of the Northeastern Linguistics Society, University of Massachusetts, Amherst, MA, October 14-16, 2016.

Landman, Fred. 2011. Count nouns, mass nouns, neat nouns, mess nouns. *The Baltic International Yearbook of Cognition, Logic and Communication* 6. 1–67.

Lima, Suzi. 2014. *The grammar of individuation and counting*. Amherst, MA: University of Massachusetts Amherst dissertation.

Scontras, Gregory. 2014. *The semantics of measurement*. Cambridge, MA: Harvard University dissertation.

Snider, Todd & Sarah E. Murray. to appear. Expressing comparison in Cheyenne. In Monica Macaulay & Margaret Noodin (eds.), *Papers of the Forty-Seventh Algonquian Conference (2015)*, Michigan State University Press.